J. Electrical Systems 20-6s (2024): 1162-1174

[1]Dr. S. Satheesh Kumar

[2]B. Mouleswararao

[3]Dr. Mahesh Maurya

[4]Dr. B.Varaprasad Rao

[5]Dr. Sreenivasulu Gogula

[6]Dr. D. Nagaraju

Modified wild Horse Herd Optimization based Bhattacharyya error constraint (BEC) based L2-norm Linear discriminant analysis (LDA) method for the sentiment analysis

Journal of Electrical Systems

Abstract: - People's opinions are analyzed via sentiment analysis in all fields. Reviews and tweets, among other formats, are used to express opinions. Irony, sarcasm, and other difficult-to-discern hidden meanings can occasionally be found in viewpoints. Artificial intelligence must be used to examine the sentiments as a result. We suggest a unique Bhattacharyya error constraint (BEC) based L2-norm linear discriminant analysis (LDA) because some of the earlier efforts lack optimization. There are some overfitting and class disparity issues in this. To address this, we used a brand-new method called Modified Wild Horse Herd Optimization (MHHO). The experiment is run to evaluate the performance of the suggested strategy and to compare it to other approaches already in use. We have used performance measures for comparison, and the results demonstrate that the suggested method successfully assesses the sentiment from the acquired dataset.

Keywords: Linear discriminant analysis; sentiment; opinion; Horse Herd optimization; GloVe; FastTex; and WordVec.

1. Introduction:

To identify and extract information, expressions, emotions, views, and other relevant data, sentiment analysis [1], also known as opinion mining, is used. The organizations turn what people think of the product or service. Expressions are classified as positive, negative, and neutral. Social media [2] enables gathering public opinion for

[1] [1]Associate Professor, Department of IT, Institute of Aeronautical Engineering

s.satheeshkumar@iare.ac.in

[2]Assoc Professor, Department of Computer Science and Engineering, Koneru lakshmaiah Education Foundation vaddeswaram guntur dt, AP

mbpalli@kluniversity.in

[3]Associate Professor and HoD, K. C. College of Engineering, Department of Computer Engineering, Mith Bunder Road, Near Hume Pipe, Kopri, Thane (E) - 400603. mahesh.maurya@kccemsr.edu.in

[4]Professor, Department of Computer Science and Engineering,RVR&JC College of Engineering,Guntur, AP.

bvpr@rvrjc.ac.in

[5]Professor and Head of the Department, Department of Computer Science and Engineering, (Data Science),Vardhaman college of Engineering, Shamshabad, Hyderabad 501218

gsrinivasulu1678@vardhaman.org

hodcsd@vardhaman.org

[6]Professor, Department of CSE,Sri Venkatesa Perumal College of Engineering and Technology, Puttur, Andhra Pradesh, India - 517583

raj2dasari@gmail.com

principal@svpcet.org

ORCID: 0000-0002-8511-1863

specific information. The word list is used to produce positive and negative messages. Before analysis, the words are developed and formulated. Text analysis, Natural language processing [3], and other techniques are utilized in sentiment analysis to measure the information. Data mining [4], machine learning [5], and artificial intelligence [6] are used to extract information. Machine learning is the method utilized for sentiment analysis. Sentiment Analysis is used to examine the positive and negative texts.

By using data from computer systems, machine learning [7] gets better in line with that. It performs some activities as a subclass of artificial intelligence (AI) [8]. Some operations can be carried out automatically. To address the difficulties encountered during the performance of deep and machine learning on the data. Linear Regression, Logistic Regression, Classification and Regression Trees, Naive Bayes, K-nearest Neighbors, Learning Vector Quantization, and Support Vector Machines are a few examples of machine learning approaches [9]. The foundation of linear regression [10] is supervised learning. It is anticipated that the target will carry out a regression task on linked variables. The categorization problems have been resolved by using logistic regression [11]. Based on a study of the algorithm, the likelihood is projected. Regression and classification trees [12] both specify the discrete value as an output value, but classification refer to the real value as the output value. Nave Bayes [13] predicts the classifier very quickly and predicts the problems more effectively. In K-nearest neighbors [14], the sign is represented by the letter "K," and the quantity of fresh data is categorized to address issues. Learning vector quantization [15] instantly selects training data and carries out various actions. To detect the line, Support Vector Machines [16] classify the N-dimensional data. The aforementioned algorithms are not sufficiently tuned, which has several drawbacks. When compared to the prior model, complexity and accuracy have decreased. While tweeting on social media, the various languages are not correctly recognized. In light of that, we put forth a brand-new Bhattacharyya error constraint (BEC) based L2-norm linear discriminant analysis (LDA) based optimization method that may be utilized to categorize feelings based on polarity, multilingualism, aspect-based reasoning, and emotions. The following is a list of our suggested word's main contributions:

➢ The Distributed Word Representation-based Domain-Specific Technique (DWR-DS) is used to map the obtained feature vectors from the acquired dataset.

➢ The mapped feature vectors are trained using the BEC-L2-norm LDA, which then classifies the comments as positive or negative depending on the many parameters described above. However, there are problems with class disparity.

➢ The MHHO algorithm, which is used for the optimized output, can get around this. Additionally, this improves the categorization output.

The remaining portion of the work is structured as follows: Section 2 lists and reviews the relevant works. In section 3, the proposed method for sentiment analysis is briefly discussed. Section 4 conducts and completely explains the experimental analysis. Section 5 provides an overview of the article.

2. Literature Survey:

A new sentiment analysis model (SLCABG) is proposed by Yang et al. [17] and is based on a sentiment lexicon. Convolutional neural networks and bidirectional gated recurrent units are used together in this technique. The length is utilized as the input to examine the efficiency, while the number of iterations is used to assess performance. To improve the suggested method, the features are assessed. As a result, the prerequisites for sentiment refinement are improved in higher necessity.

Rehman et al. [18] have presented a hybrid Convolutional Neural Network-Long Short-Term Memory to address the sentiment analysis difficulties (CNN-LSTM). To construct features for local data, the Word2Vec model is employed. a huge group of words culled from stored semantic information. Both convolutional and long short-term memories are compared when the model is integrated to extract features and varying lengths. However, the level of intricacy needs to be raised.

Deep transfer learning to recognize In Russia, Smetanin et al. [19] suggest a baseline for attitude analysis. To identify the language and carry out transfer learning, the model is supported. The text data is automatically determined. There are solid baselines for transferring learning categorization. Different language models are supported by the discovered sentiment analysis datasets. However, these models specify the datasets.

Highlighting the attributes of aspect-level sentiment detection has been proven by Nandal et al. [20]. Customer reviews on Amazon are evaluated for the identified data. The following pre-processing steps are utilized to extract the dataset: casing, stop-word removal, tokenization, and stemming. The sentiment analysis is used to build the strategy and is reviewed. Users can benefit from understanding the software's product. So the issues to be resolved are spam and sarcasm.

To assess feelings A method dubbed "communalizing user data" is suggested by Chakraborty et al [21]. Social networks can be performed using Social Network Analysis (SNA). Retweets, likes, and shares allow the social network to gauge the product's quality. A significant factor is raising the standard of social media and networks. Recognizing the user's opinion is quite effective. As a result, future health issues should be addressed.

Bidirectional Encoder Representations from Transformers (BERT) is a paradigm that Singh et al. [22] presented for sentiment analysis. There are two datasets in it. The BERT model is connected to the classification of emotions. The social network moves very quickly, and the classification of emotions is more accurate. Therefore, the complexity needs to be raised.

3. Proposed machine learning-based approach for sentiment analysis

This section explains the intended work. This makes use of the mapping of feature vectors with the assistance of the DWR-DS technique, and after that, the innovative Bhattacharyya error constraint (BEC) based L2-norm LDA does the sentiment classification from the reviews and tweets. However, there are concerns about the class imbalance that can be resolved with the MHHO method. By adjusting the parameters, this accepted algorithm can be utilized to improve the output that is categorized. Figure 1 explains the schematic overlay of our suggested strategy.

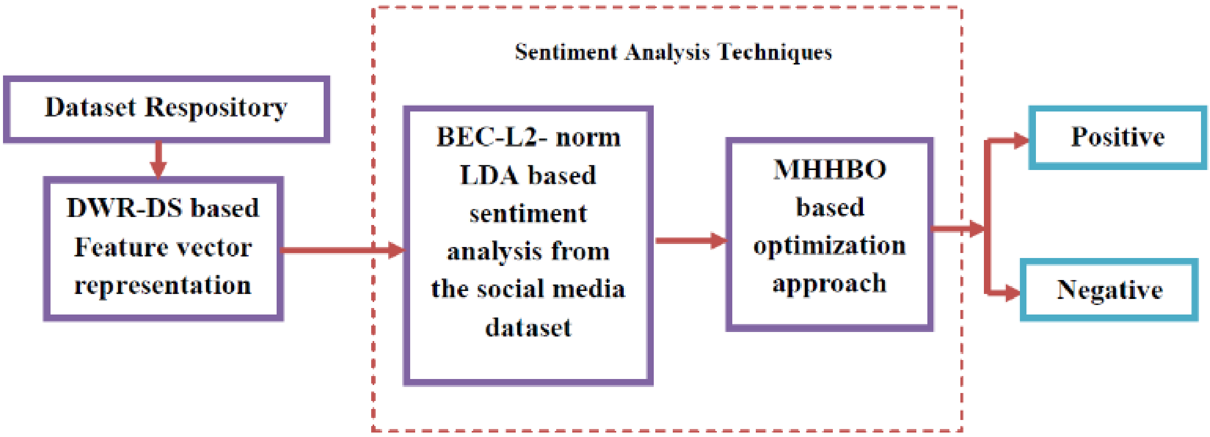

Fig 1: Schematic overlay of the proposed approach

3.1 DWR-DS

Using the broad term distributed word representation [23], the feature vectors from the tweets or reviews are mapped. The feature vectors in the entry indicate the words' underlying meaning. Additionally, it illustrates the connection between semantic and syntactic properties. Our suggested method precisely extracts the necessary information, which improves the accuracy of the social media network's sentiment analysis. Semantic and contextual advertising is the major terminology employed in the sentiment analysis of social media networks. Modern works categorize the texts and their various judgments in most cases [24]. It is necessary to examine the sentiment using semantic and contextual factors. The most popular feature extraction techniques include GloVe [25], Word2vec [26], and FastText [27].

➢ **Word2Vec**: The pertinent words for the sentences that are broken apart can be found using this technique. The learning procedures are carried out using deep learning methodologies. The vector accurately arranges the words in the correct sequence. The majority of researchers mainly employ the Skip-Gram approach [28]. Consequently, the Skip-Gram technique's objective function can be expressed as follows:

$$\frac{1}{w}\sum_{W=1}^{w}\sum_{-s\le j\le s,\, j\ne 0}\log(R(d_{t+j}|d_W)) \tag{1}$$

The total number of training words from the dataset is denoted as w. The size of the training context is indicated as s, d is the training word, the center word in the training context can be denoted as d_W . If the value of s possesses maximum value then it denotes the maximum accuracy.

➢ **FastText:** This method primarily makes use of Facebook's AI Research lab's (FAIR) [19] word and representation library. This methodology, also known as the skip-gram approach, has the following objective function:

$$f(d,s)=\sum_{h\in H_d} z_h^W l_s \tag{2}$$

The scoring function can be denoted as;*f*. *H* denotes the size of the n-grams. The values of H_d fall under the range of 1 to *H* with *d* words. Each n-gram *h's* vector representation can be denoted as z_h. The context vector can be represented as l_s. Hence this is more reliable than the previous one.

➢ ***GloVe*:** The overall vector representation looks like this. Word2Vec and GloVe are two tools that can successfully extract the local window context and assess the global comprehensive co-occurrence statistics of words.

The sentiment analysis makes use of polarity, synonymous terms, emotions, and other functions. It is challenging to recognize these patterns; hence we suggested a novel BEC-L2-norm LDA-based technique to address these problems. The next section provides more details about this.

3.2 Linear Discriminant analysis

The projected data samples with more discriminative data are used to identify the low-dimensionality space data using the LDA [30]. Data dimensionality has been reduced with the aid of a linear transformation matrix. The linear transformation matrix is given as, $M\in P^{n\times r}, r\le n$. The between-class scattering matrix X_m and scatter matrix within-class X_c can be formulated as,

$$X_m=\frac{1}{Q}\sum_{i=1}^{a}Q_i(\bar{e}_i-\bar{e})(\bar{e}_i-\bar{e})^T \tag{3}$$

and

$$X_c=\frac{1}{Q}\sum_{i=1}^{a}\sum_{x=1}^{Q_i}(e_x^i-\bar{e}_i)(e_x^i-\bar{e}_i)^T \tag{4}$$

The optimization issues are subjugated by using the following equations:

$$\max_c \frac{Tr(c^T X_m c)}{Tr(c^T X_c c)} \tag{5}$$

The trace operation used in the equation is denoted as $Tr(.)$ and the transposed matrix can be indicated as $(.)^T$. From the generalized issues $X_m c = \gamma X_c c,\ \gamma \neq 0$, the optimal solutions can be acquired as $C = (c_1,....,c_r)$. The value of C is stated as the maximized eigenvalue d from the $X_c^{-1} X_m$ if the nonsingular value is given as X_c.

3.3 BCE-based L2-norm LDA

The limitations of the Bhattacharyya error have been reduced in the suggested method, enhancing the separation of L2 norm LDA-based between class scatterings. This decreases the scattering distance within and is computed as the weighted sum of pairwise distances of class means. Consequently, it can be stated as,

$$X_m^M = \frac{1}{Q}\sum_{i<j}^{Q}\sqrt{Q_i Q_j}\left(\bar{e}_i - \bar{e}_j\right)\left(\bar{e}_i - \bar{e}_j\right)^T \tag{6}$$

and,

$$X_c^B = \sum_{i=1}^{a}\sum_{x=1}^{Q_i}\left(\bar{e}_x^i - \bar{e}_i\right)\left(\bar{e}_x^i - \bar{e}_i\right)^T \tag{7}$$

The BCE based L2-norm LDA can be estimated as,

$$\begin{aligned}&\min_c - Tr\left(c^T X_m^M c\right) + \Delta . Tr\left(c^T X_c^C c\right)\\ &s.t. c^T c = I\end{aligned} \tag{8}$$

Here, $c \in P^{q\times d}$, where $d \leq q$ and $\Delta = \frac{1}{4}\sum_{i<j}^{a}\sqrt{R_i R_j}\left\|\bar{e}_i - \bar{e}_j\right\|_2^2, R_i = \frac{Q_i}{Q}, R_j = \frac{Q_j}{Q}$.

The interval between the ith class means and each class means is higher in the low dimensionality area if the objective function is minimized. $\frac{1}{Q}\sqrt{Q_i Q_j}$ represents the coefficient of the weighted distance between the ith and jth class. The coefficient of within-class scatter is denoted as Δ. This also ensures the reduction in error constraints. The orthonormality property of discriminant directions is provided by the limited $c^T c = I$. Our proposed L2 norm LDA is used to solve the issues via the utilization of the eigenvalue of the decomposition issues as shown below,

$$\begin{aligned}&\min_c - Tr\left(c^T X c\right)\\ &s.t. c^T c = I\end{aligned} \tag{9}$$

Here, $X = -X_a^A + \Delta . X_c^A$ and can be used to classify the sentiment based on the reviews or tweets. There are, however, certain overfitting and class disparity problems. Therefore, we suggested an MHHO algorithm to resolve those problems. This improves the suggested method's classification accuracy based on the negative and positive classes. The following section explains.

3.4 Modified wild Horse herd optimization (MHHO) algorithm

The algorithm HHO has evolved depending on the type of horse pattern. There are six traits of horses included in it, including the grazing mechanism, hierarchical mechanism, amiability mechanism, and wandering mechanism [31]. The equation for the horse's motion during each round is as follows:

$$B_n^{Ro\cdot Y} = \overrightarrow{W}_n^{Ro\cdot Y} + B_n^{(Ro-1)\cdot Y}, \quad Y = \theta, \rho, \psi, \phi \tag{10}$$

The location of the nth horse is represented as $B_n^{Ro\cdot Y}$. The velocity of the horse while moving is denoted as $\overrightarrow{W}_n^{Ro\cdot Y}$. The characters of horses vary with age and surroundings.

- **Grazing (GR):**

The horse is a species of grazing animal that consumes green plants and grasses. It can be written mathematically as,

$$\overrightarrow{GR}_n^{Ro.Y} = f_{Ro}\left(\tilde{UR} + \beta \tilde{LR}\right)\left[B_n^{Ro-1}\right] \tag{11}$$

$$f_n^{Ro\cdot Y} = f_n^{(Ro-1)\cdot Y} \times V_f \tag{12}$$

The parameter used to indicate the movement of the jth horse is $f_n^{(Ro-1)\cdot Y}$. The lower and upper constraints are given as UR and LR correspondingly.

- **Hierarchy (H):**

The hierarchical behavior of the horses is determined by the tendency factors and can be represented as T_F. The middle-age horses ρ, ψ can be expressed as,

$$\overrightarrow{H}_n^{Ro\cdot Y} = h_n^{Ro\cdot Y}\left[A_*^{(Ro-1)} - A_n^{(Ro-1)}\right] \tag{13}$$

$$h_n^{Ro\cdot Y} = h_n^{(Ro-1)\cdot Y} \times V_h \tag{14}$$

The location velocity of the best horse is represented as $\overrightarrow{H}_n^{Ro\cdot Y}$ and its respective position is indicated as $A_*^{(Ro-1)}$.

- **Sociability (SL):**

The average size is designated as HL, and the horse's security is compromised by predators' hunting habits.

$$\overrightarrow{SL}_n^{Ro\cdot Y} = HL_n^{Ro\cdot Y}\left[\left(\frac{1}{L}\sum_{k=1}^{L} a_k^{(Ro-1)}\right) - A_n^{(Ro-1)}\right] \tag{15}$$

$$HL_n^{Ro\cdot Y} = HL_n^{(Ro-1)\cdot Y} \times V_{HL} \tag{16}$$

The social movement of the jth horse is represented as $HL_n^{(Ro-1)\cdot Y}$. The respective orientation of the horse in the round is given as $HL_n^{(Ro-1)\cdot Y}$. With the reduction factor, V_{HL} the value of the iteration is reduced to $HL_n^{Ro\cdot Y}$.

- **Defense scheme (DF)*:***

The defense system can be represented as DF. Mathematically it can be expressed as,

$$\overrightarrow{DF}_n^{Ro\cdot Y} = -d_n^{Ro\cdot Y}\left[\left(\frac{1}{NL}\sum_{k=1}^{NL} A_k^{(Ro-1)}\right) - A^{(Ro-1)}\right] \tag{17}$$

$$d_n^{Ro\cdot Y} = -d_n^{(Ro-1)} \times V_d \tag{18}$$

The escape vector of the average horses is denoted as $\overrightarrow{DF}_n^{Ro\cdot Y}$. The worst position of the horse with the number of the NL is given as $\vec{A}$.

4. Result and Discussion:

The dataset description and the experimental settings are covered in this section. In addition, a several assessment measures have been subjected to experimental evaluation, as illustrated below. Designers used the most popular Python programming framework, NLTK (Natural Language Tool Kit) [32, 33], to assess the sentiment in tweets. There are also user-friendly interfaces for more than 50 corpora and lexical resources, including WordNet, as well as text processing libraries for parsing, tokenization, stemming, semantic reasoning categorization, and tagging [34–36]. We employed the clustering, regression, and classification techniques found in the Python Scikit-learn toolkit.

4.1 Explanation of datasets:

In this work, the proposed approach is evaluated using the Apple Twitter sentiment document (71585 reviews) and the Amazon fine food reviews (568,454 reviews) datasets. The tweets from the two datasets were combined, and we produced three pairs of randomly shuffled tweets that we dubbed AST1, AST2, and AST3 [37, 38].

4.2 Performance measures:

Performance indicators like recall, accuracy, and F1-Score are used for evaluating our suggested technique.

❖ *Accuracy*

The phrase "accuracy or precision" relates to how often a sentiment analysis was correct. If a couple of these tonalities are accurately scored, for example, the accuracy of papers with the tone can be assessed.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{1}$$

❖ *Recall*

The recall is a measurement of the frequency with which emotional documents were classified as such. i.e., what neutrality means to the system.

$$\mathrm{Re}call = \frac{T_P}{T_P + F_N} \tag{2}$$

❖ *F1 Score:*

This can be described as a combination of accuracy and recall. Most usually, the F1 score falls between 0.0 and 1.0.

$$F1 - score = 2 \times \frac{\mathrm{Re}call * \mathrm{Pr}ecision}{\mathrm{Re}call + \mathrm{Pr}ecision} \tag{3}$$

4.3 Performance analysis:

Every strategy was paired with the visual words to evaluate the proposed work performance, and it was then compared to traditional SNA, CNN-LSTM, and BERT techniques. In the comparison research, individual pre-trained distributed word representations along similar structures were used. Additionally, accuracy and F1 score are estimated using baseline classifiers. Figures 2 and 3 present the outcomes, respectively.

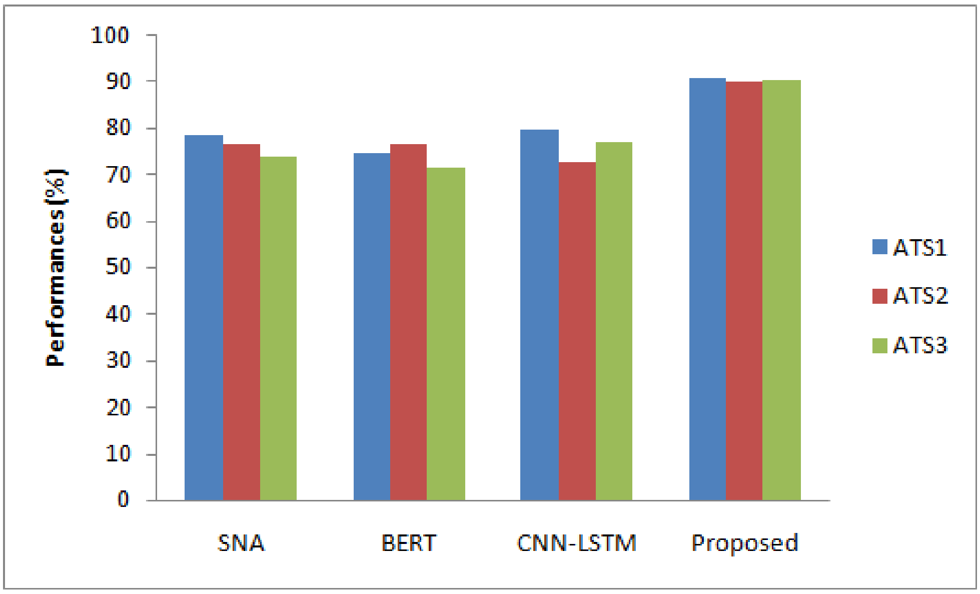

Figure 2: Analysis of accuracy based on 3 arbitrarily jumbled-up tweets

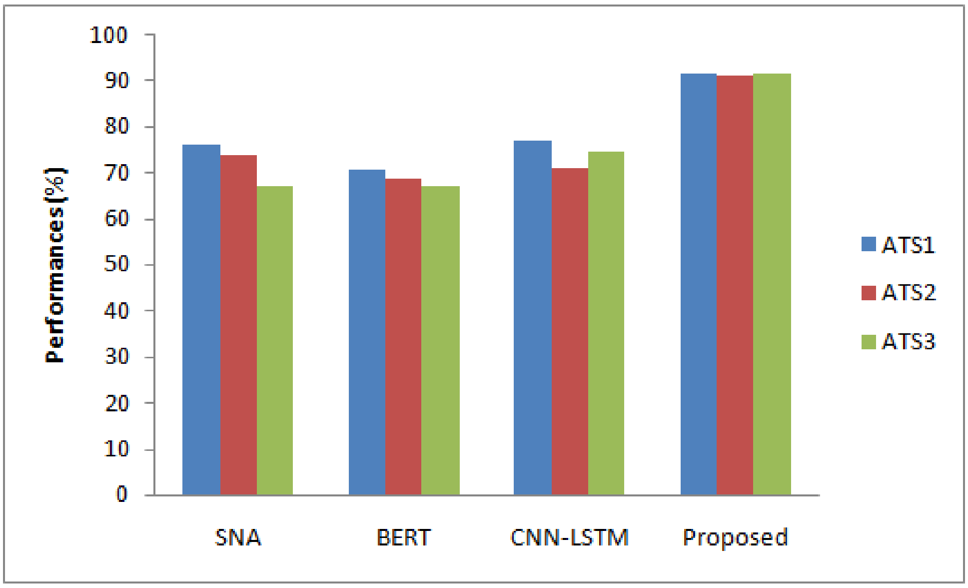

Figure 3: Analysis ofF1 Score based on 3 arbitrarily jumbled-up tweets

To assess the planned work's overall performance, certain carefully chosen data aspects have previously been examined. The proposed effort to existing studies for the top 50 data features sets has been evaluated, as shown in Figure 4. In this instance, the suggested method obtains 94 percent accuracy and 94.67 percent F1 scores. This leads to a balance between exploitation and investigation in the proposed study, as well as precise categorization based on a range of emotions.

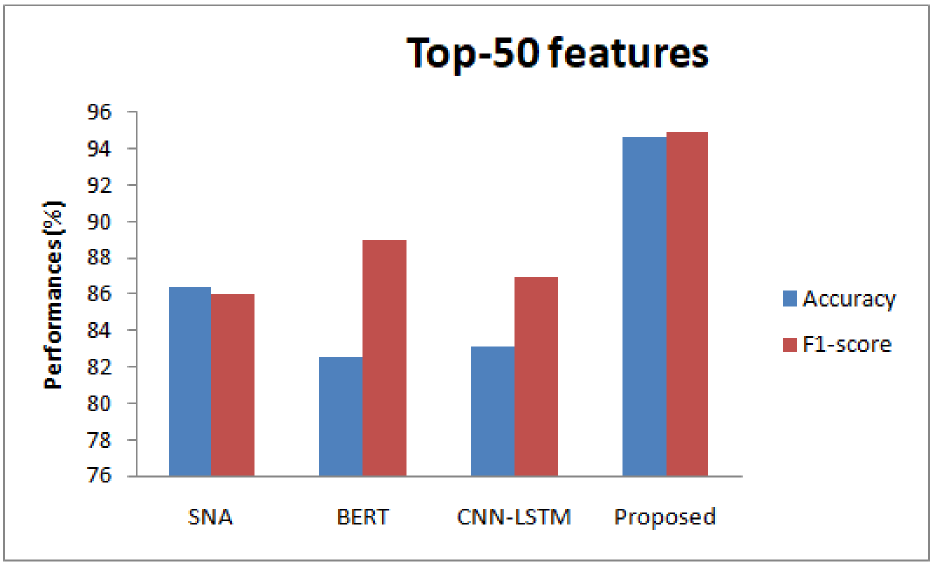

Figure 4: Evaluation of accuracy and F-score with respect to 50 features

Figure 5 shows the overall evaluations for the top 100 feature data sets for both the proposed and current work. SNA, BERT, and CNN-LSTM are examples of existing techniques that are utilized to validate the effectiveness of the suggested strategy. The proposed method has an accuracy of 87.45 percent for the F1 score of 0.96.

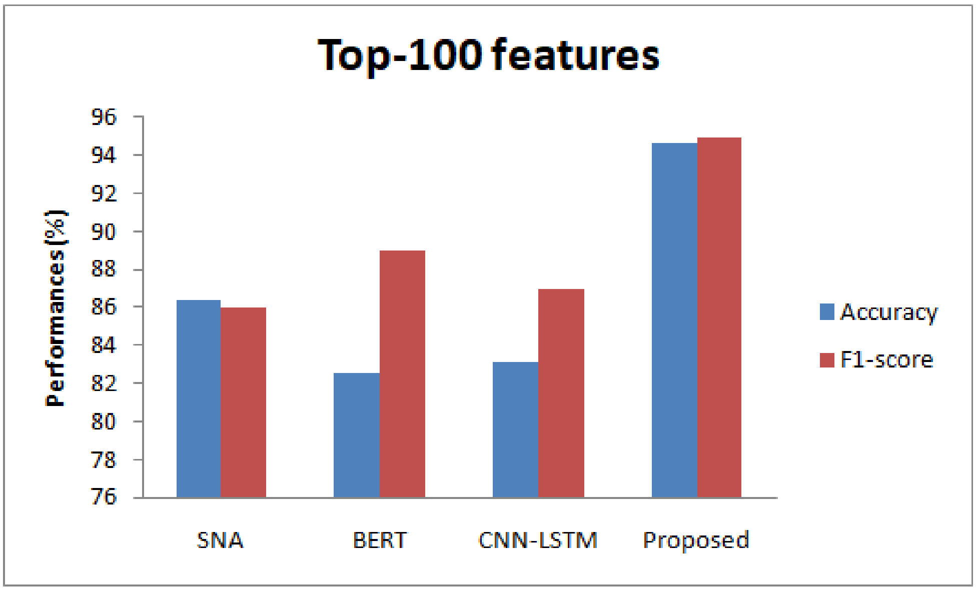

Figure 5: Evaluation of accuracy and F-score with respect to 100 features

Figure 6 compares the overall performance of the proposed work to that of earlier systems in terms of the top 200 chosen attributes. The recommended strategy had an F1 score of 0.91 and an accuracy of 85.76 percent according to the investigative analysis. The accuracy has decreased as the value of feature sets has increased. The current approach's accuracy has likewise declined.

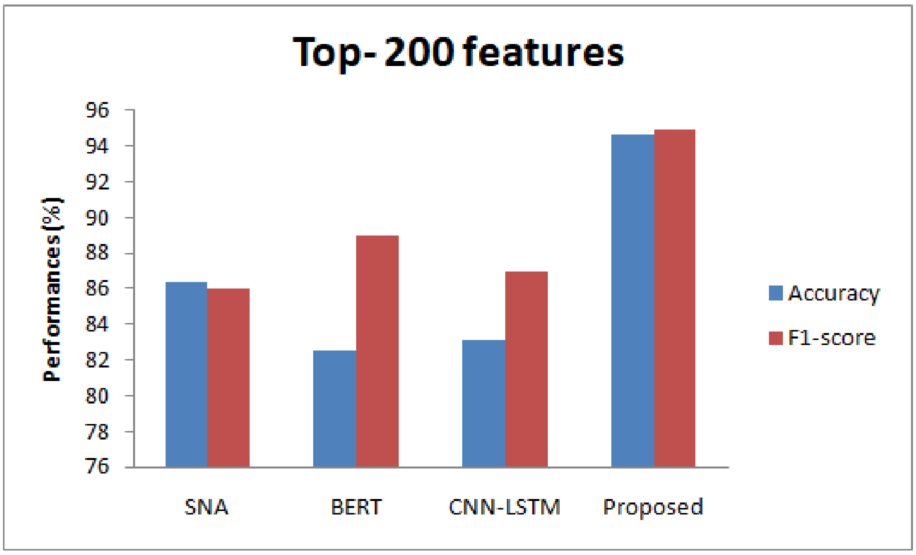

Figure 6: Evaluation of accuracy and F-score with respect to 200 features

Figure 7 shows the overall effectiveness of the suggested methodology compared to all other methods for the whole feature dataset. Although the accuracy improved slightly over the previous research, it was still inferior to the small data sets. Overall, the proposal beats state-of-the-art approaches in terms of dataset features. Therefore, the proposed work outperforms all other efforts in terms of all dataset attributes. The accuracy did, however, decline as the number of characteristics did.

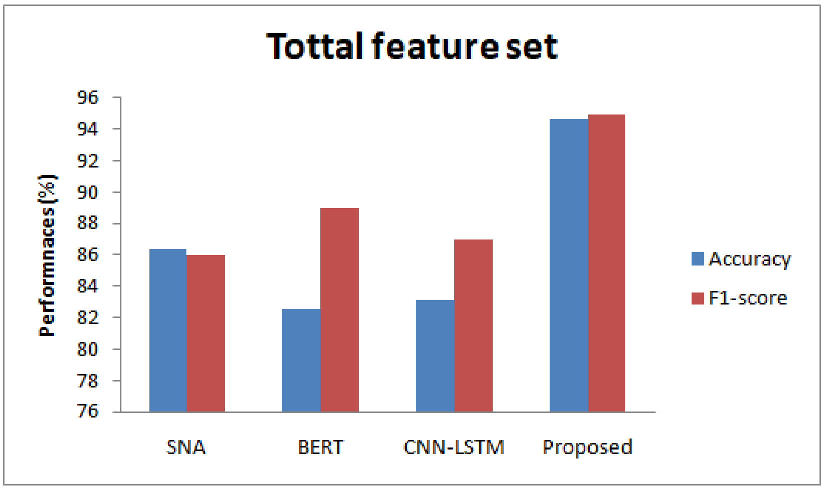

Figure 7: Evaluation of accuracy and F-score with respect to whole feature sets

Table 1 provides an illustration of a state-of-the-art evaluation based on sentiment analysis. In this inquiry, the standard SNA, CNN-LSTM, and BERT methodologies are combined with the suggested methodology. Compared to other approaches now in use, the proposed method gives good outcomes.

Table 1: State-of-art evaluation with respect to sentiment analysis

Example	**SNA**	**BERT**	**CNN-LSTM**	**Proposed**
I receive a message that my storage is nearly full approximately every five minutes.	Negative	Positive	Negative	Negative
Thank you for using xtracheckin to check in at the Upper Westside location. However, why	Positive	Positive	Positive	Negative

are appointments running approximately 50 minutes behind schedule?				
When the items arrived, they were marked as "Jumbo Salted Peanuts," however they were actually little and unseasoned. I'm not sure if the dealer accidentally named the items "Jumbo" or if that was their intention.	Negative	Negative	Positive	Negative

5. Conclusion

The analysis of the sentiment from the opinions expressed by the people is based on the BEC-L2-norm LDA technique in this article. Additionally, there are certain class imbalances and overfitting difficulties with the suggested approach. The use of the MHHO algorithm has decreased this. The characteristics were retrieved using the DWR-DS technique prior to classification. The characteristics from the dataset were successfully retrieved in this way. This technique efficiently examines the sentiment from several angles. The comparison analysis and experimental analysis show that the suggested approach is more effective at analyzing sentiment than the other approaches.

Compliance with Ethical Standards

Conflict of interest

The authors declare that they have no conflict of interest.

Human and Animal Rights

This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed Consent

Informed consent does not apply as this was a retrospective review with no identifying patient information.

Funding: Not applicable

Conflicts of interest Statement: Not applicable

Consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material:

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Code availability: Not applicable

References

[1] Shaukat, Z., Zulfiqar, A.A., Xiao, C., Azeem, M. and Mahmood, T., 2020. Sentiment analysis on IMDB using lexicon and neural networks. *SN Applied Sciences*, *2*(2), pp.1-10.

[2] Kaufhold, M.A., Rupp, N., Reuter, C. and Habdank, M., 2020. Mitigating information overload in social media during conflicts and crises: design and evaluation of a cross-platform alerting system. *Behaviour & Information Technology*, *39*(3), pp.319-342.

[3] Rajput, A., 2020. Natural language processing, sentiment analysis, and clinical analytics. In *Innovation in Health Informatics* (pp. 79-97). Academic Press.

[4] Dogan, A. and Birant, D., 2021. Machine learning and data mining in manufacturing. *Expert Systems with Applications*, *166*, p.114060.

[5] Sarker, I.H., 2021. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), pp.1-21.

[6] Ruan, Q., Wang, Z., Zhou, Y. and Lv, D., 2020. A new investor sentiment indicator (ISI) based on artificial intelligence: A powerful return predictor in China. *Economic Modelling*, *88*, pp.47-58.

[7] Hew, K.F., Hu, X., Qiao, C. and Tang, Y., 2020. What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education*, *145*, p.103724.

[8] Ferreira, F.G., Gandomi, A.H. and Cardoso, R.T., 2021. Artificial intelligence applied to stock market trading: a review. *IEEE Access*, *9*, pp.30898-30917.

[9] Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A. and Choi, G.S., 2021. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *Plos one*, *16*(2), p.e0245909.

[10] Dong, J., Chen, Y., Gu, A., Chen, J., Li, L., Chen, Q., Li, S. and Xun, Q., 2020. Potential Trend for Online Shopping Data Based on the Linear Regression and Sentiment Analysis. *Mathematical Problems in Engineering*, *2020*.

[11] Singh, R. and Singh, R., 2021. Applications of sentiment analysis and machine learning techniques in disease outbreak prediction–A review. *Materials Today: Proceedings*.

[12] Madasu, A. and Elango, S., 2020. Efficient feature selection techniques for sentiment analysis. *Multimedia Tools and Applications*, *79*(9), pp.6313-6335.

[13] Li, Z., Li, R. and Jin, G., 2020. Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary. *Ieee Access*, *8*, pp.75073-75084.

[14] Isnain, A.R., Supriyanto, J. and Kharisma, M.P., Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), pp.121-130.

[15] Pourghasemi, H.R., Gayen, A., Lasaponara, R. and Tiefenbacher, J.P., 2020. Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. *Environmental research*, *184*, p.109321.

[16] Xia, H., Yang, Y., Pan, X., Zhang, Z. and An, W., 2020. Sentiment analysis for online reviews using conditional random fields and support vector machines. *Electronic Commerce Research*, *20*(2), pp.343-360.

[17] Yang, L., Li, Y., Wang, J. and Sherratt, R.S., 2020. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*, *8*, pp.23522-23530.

[18] Rehman, A.U., Malik, A.K., Raza, B. and Ali, W., 2019. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications*, *78*(18), pp.26597-26613.

[19] Smetanin, S. and Komarov, M., 2021. Deep transfer learning baselines for sentiment analysis in Russian. *Information Processing & Management*, *58*(3), p.102484.

[20] Nandal, N., Tanwar, R. and Pruthi, J., 2020. Machine learning based aspect-level sentiment analysis for Amazon products. *Spatial Information Research*, *28*(5), pp.601-607.

[21] Chakraborty, K., Bhattacharyya, S. and Bag, R., 2020. A survey of sentiment analysis from social media data. *IEEE Transactions on Computational Social Systems*, *7*(2), pp.450-464.

[22] Singh, M., Jakhar, A.K. and Pandey, S., 2021. Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining*, *11*(1), pp.1-11.

[23] Le-Khac, P.H., Healy, G. and Smeaton, A.F., 2020. Contrastive representation learning: A framework and review. *IEEE Access*, *8*, pp.193907-193934.

[24] Nemes, L. and Kiss, A., 2021. Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, *5*(1), pp.1-15.

[25] Wen, F., Sun, Z., He, T., Shi, Q., Zhu, M., Zhang, Z., Li, L., Zhang, T. and Lee, C., 2020. Machine learning glove using self-powered conductive superhydrophobic triboelectric textile for gesture recognition in VR/AR applications. *Advanced science*, *7*(14), p.2000261.

[26] Kim, S., Park, H. and Lee, J., 2020. Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis. *Expert Systems with Applications*, *152*, p.113401.

[27] Al Badawi, A., Hoang, L., Mun, C.F., Laine, K. and Aung, K.M.M., 2020. Privft: Private and fast text classification with homomorphic encryption. *IEEE Access*, *8*, pp.226544-226556..

[28] Mohan, A., Prabha, G. and V., A. 2023. Multi Sensor System and Automatic Shutters for Bridge- An Approach. International Journal of Intelligent Systems and Applications in Engineering. 11, 4s (Feb. 2023), 278–281.

[29] Prabha , G. , Mohan, A. , Kumar, R.D. and Velrajkumar, G. 2023. Computational Analogies of Polyvinyl Alcohol Fibres Processed Intellgent Systems with Ferrocement Slabs. International Journal of Intelligent Systems and Applications in Engineering. 11, 4s (Feb. 2023), 313–321.

[30] Study On Structural Behaviour Of Ductile High-Performance Concrete Under Impact And Penetration Loads, Lavanayaprabha, S. Mohan, A. Velrajkumar, G., Mohammedharoonzubair, A. Journal of Environmental Protection and Ecology., 2022, 23(6), pp. 2380–2388..

[31] Mohan, A., & K, S. . (2023). Computational Technologies in Geopolymer Concrete by Partial Replacement of C&D Waste. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 282–292.

[32] Mohan, A., Dinesh Kumar, R. and J., S. 2023. Simulation for Modified Bitumen Incorporated with Crumb Rubber Waste for Flexible Pavement. International Journal of Intelligent Systems and Applications in Engineering. 11, 4s (Feb. 2023), 56–60.

[33] R.Gopalakrishnan, Mohan, "Characterisation on Toughness Property of Self-Compacting Fibre Reinforced Concrete", Journal of Environmental Protection and Ecology 21, No 6, 2153–2163 (2020).

[34] Bankapur, S. and Patil, N., 2020. An enhanced protein fold recognition for low similarity datasets using convolutional and skip-gram features with deep neural network. *IEEE Transactions on Nanobioscience*, *20*(1), pp.42-49.

[35] Guess, A., Nagler, J. and Tucker, J., 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science advances*, *5*(1), p.eaau4586.

[36] Guo, Y.R., Bai, Y.Q., Li, C.N., Shao, Y.H., Ye, Y.F. and Jiang, C.Z., 2021. Reverse nearest neighbors Bhattacharyya bound linear discriminant analysis for multimodal classification. *Engineering Applications of Artificial Intelligence*, *97*, p.104033.

[37] Naruei, I. and Keynia, F., 2021. Wild horse optimizer: A new meta-heuristic algorithm for solving engineering optimization problems. *Engineering with Computers*, pp.1-32.

[38] Hassan, M.K., Hudaefi, F.A. and Caraka, R.E., 2021. Mining netizen's opinion on cryptocurrency: Sentiment analysis of Twitter data. Studies in Economics and Finance.

[39] Antonakaki, D., Fragopoulou, P. and Ioannidis, S., 2021. A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. Expert Systems with Applications, 164, p.114006.

[40] Kausar, M.A., Soosaimanickam, A. and Nasar, M., 2021. Public sentiment analysis on Twitter data during COVID-19 outbreak. Int. J. Adv. Comput. Sci. Appl, 12(2), pp.415-422.

[41] Kharde, V. and Sonawane, P., 2016. Sentiment analysis of twitter data: a survey of techniques. arXiv preprint arXiv:1601.06971.

[42] Gautam, G. and Yadav, D., 2014, August. Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In 2014 Seventh international conference on contemporary computing (IC3) (pp. 437-442). IEEE.

[43] Liao, S., Wang, J., Yu, R., Sato, K. and Cheng, Z., 2017. CNN for situations understanding based on sentiment analysis of twitter data. Procedia computer science, 111, pp.376-381.

[44] Pagolu, V.S., Reddy, K.N., Panda, G. and Majhi, B., 2016, October. Sentiment analysis of Twitter data for predicting stock market movements. In 2016 international conference on signal processing, communication, power and embedded system (SCOPES) (pp. 1345-1350). IEEE.